Biometric Identification Based on Frequency Analysis of Cardiac Sounds

Francesco Beritelli, *Member, IEEE*, and Salvatore Serrano

***Abstract*—The performance of traditional biometric identification systems is, as yet, unsatisfactory in certain applications. For this reason, other physiological or behavioral characteristics have recently been considered, using new electrical or physical signals linked to a person's vital signs. This paper examines the biometric characteristics of PhonoCardioGram (PCG) signals from cardiac auscultation. The idea is that PCG signals have specific individual characteristics that can be taken into consideration as a physiological sign used in a biometric system. More specifically, the paper proposes a preliminary study related to the identification of individuals via frequency analysis of cardiac sounds. The results, obtained using a database containing several heart sound recordings from 20 different people, confirm the biometric properties of PCG signals, which can thus be included among the physiological signs used by an automatic identification system.**

***Index Terms*—Automatic detection of cardiac sounds, biometric identification, frequency analysis of PhonoCardioGram (PCG) signals.**

I. Introduction

THE recognition and identification of subjects using biometric technologies usually aims to implement security systems in which the "immediate and certain" recognition of a person is a primary objective [1], [2]. In this perspective, solutions based on the use of biometric technology supplement traditional solutions based on the use of passwords, PIN codes, smart cards, etc., with unique individual characteristics. Various market sectors require solutions in which biometric technology will guarantee the "safe" recognition of individuals [3], [4]. The solutions most frequently requested refer to:

- physical access control—buildings, parking lots, offices, hospitals, and prisons;
- logical access control—PCs, computer networks (intranet), mobile phones, e-mail, confidential data (e.g., clinical files), and banking services;
- surveillance—sports events, shopping centers, concerts, buildings, and museums.

Biometric technology is based on well-defined physiological or behavioral characteristics. This technology, used together with traditional solutions, provides a highly efficient level of security in cases where it is essential to obtain certain identification, even in unfavorable environmental conditions. The individual physical characteristics used in a biometric system cannot be forgotten as they are intrinsically individual properties. It is also true, however, that some of them may be "lost" following serious trauma, or deliberately altered, by plastic surgery for example. In addition, due to various (physical or environmental) factors, the result of a biometric recognition process is never completely accurate and performance in terms of the false reject rate (FRR) and false accept rate (FAR) is not yet satisfactory in some fields of applications [1]. For this reason, other physiological or behavioral characteristics have recently been taken into consideration, looking at the human body as a constant source of new "biological documents" (i.e., implementing automatic biometric recognition algorithms which use new electrical or physical signals linked to an individual's vital signs). This paper focuses, in particular, on the biometric characteristics of PhonoCardioGram (PCG) signals from cardiac auscultation [5]. The recent introduction of electronic stethoscopes, which provide excellent-quality digital PCG signals, has aroused considerable interest in techniques for the automatic analysis of PCG signals, which contain a large amount of information about the state of an individual's cardiocirculatory system. The idea is that PCG signals have specific individual characteristics that can be taken into consideration as a physiological sign used in a biometric system. This paper proposes a study related to algorithms for human identification via frequency analysis of an individual's cardiac sounds. The results obtained confirm the biometric properties of PCG signals, which can thus be included among the physiological signs used by an automatic identification system. More specifically, this paper shows the robustness of both the segmentation and matching algorithms by analyzing their performance first on a small number of PCG sequences taken from a subset of a cardiosource database [6] and then on a larger database specifically created via a series of PCG sequence recordings using an electronic stethoscope. The test results show that the proposed algorithms have a good capacity to correctly recognize the true or false identity of a person. The paper is organized as follows. Section II describes the phases of the biometric identification process. Section III presents the detection algorithm used for automatic identification of the main cardiac sounds S1 and S2 and the matching algorithm proposed for the identification phase based on frequency analysis of the cardiac sounds S1 and S2. A study on robustness to white Gaussian noise is also presented. Section IV describes

Manuscript received November 1, 2006; revised April 10, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Josef Bigun.

The authors are with the Dipartimento di Ingegneria Informatica e delle Telecomunicazioni, Università degli Studi di Catania, Catania 95125, Italy (e-mail: beritelli@diit.unict.it; sserrano@diit-unict.it).

Digital Object Identifier 10.1109/TIFS.2007.902922

the results obtained in terms of intra/interperson distance. Finally, Section V concludes the paper.

II. Biometric Identification

Biometric technologies are based on the use of individual characteristics for the recognition or identification of an individual [7]. They are divided into two areas:

1) physiological characteristics (unique and unvarying), which include the geometry of the hand and the palm print, fingerprints, retina, or iris image, and the (geometrical) features of the face;
2) behavioral characteristics (unique but varying), which include signature, way of walking, voice (the latter also belongs to the previous group), and keyboard typing style.

A feature common to all biometric technologies is the capability of human recognition from biometric data. This consists of a series of basic processes:

1) acquisition and storage of reference biometric data acquired by means of sensors (optical, ultrasonic, thermal, etc.);
2) acquisition of new biometric data at the start of a recognition process, for comparison with the reference data;
3) determination of the correspondence of the newly acquired data to the stored reference to determine whether they both could have been generated by the same person.

In the first phase (enrollment), a sample is acquired, which might be converted to a template, a model, or left unprocessed. This is accomplished in a controlled environment so as to guarantee the security of the original and the properties of the biometric print. The second phase (matching) is activated whenever it is necessary to verify or identify a print. Authentication or verification is performed by comparing the acquired print with a reference. Identification is performed by searching in an archive for a compatible template. They therefore serve to identify a person whose identity is not necessarily known *a priori*. The implementation of a biometric solution requires recognition threshold values to be assigned. Due to various physical and environmental factors, in fact, the result of a biometric identification process is never completely certain. The definition of the recognition threshold establishes the boundary between acceptance or rejection of a request. In this way, it is possible to set the security level on which the false reject rate (FRR) depends on (i.e., the number of times the system does not recognize a sample as coming from the same individual who produced the reference) and the false accept rate (FAR) (i.e., the number of times the system incorrectly matches a sample from one person to a reference from another). The FRR and FAR are generally considered as indexes of the biometric system's performance.

III. Cardiac Sound Extraction and Analysis

A. General Information About Heart Sounds

The mechanisms that generate cardiac sounds are complex and there is, as yet, no general consensus as to the contribution made by various mechanical cardiac events to the formation of the single components. The acoustic phenomena include:

- sounds: brief vibrations caused by the closure of the valves and tensing of the cardiac muscle;
- murmurs: caused by turbulence in the blood flow through narrow cardiac valves or reflow through the atrioventricular valves due to congenital or acquired defects.

The two loudest sounds that can be heard in any individual are the first and second sounds, referred to as S1 and S2 [5]. S1 has a typical duration of about 150 ms and S2 usually lasts about 120 ms. The first is associated with the closure of the mitral tricuspid valve during isovolumetric contraction of the ventricles. The second is associated with the closure of the aortic pulmonary valve during isovolumetric relaxation, when the ventricles end ejection and start the diastole.

Rarely, a third and a fourth heart sound (S3, S4) might be heard. The third heart sound or protodiastolic sound is not of valvular origin, as it occurs at the beginning of diastole just after S2. This sound occurs when the left ventricle is not very compliant, and at the beginning of diastole, the rush of blood into the left ventricle causes vibration of the valve leaflets and the chordae tendinae. The third heart sound is normal in children and young adults, but disappears before middle age. Abnormal re-emergence of this sound late in life indicates a pathological state, often a sign of a failing left ventricle as in congestive heart failure. This sound is called a protodiastolic gallop, a type of gallop rhythm. The rare fourth heart sound S4 is sometimes audible in healthy children, but when audible in an adult, it is called a presystolic gallop. This gallop is a sign of a pathologic state, usually a failing left ventricle. This sound occurs just after atrial contraction. The combined presence of S3 and S4 is a quadruple gallop. At rapid heart rates, S3 and S4 may merge to produce a summation gallop.

B. Cardiosource Database

The first phase was performed using a subset of a database called "heart songs" [6], containing heart sounds from people suffering from various types of cardiac pathology. In order to study the spectral characteristics of heart sounds and, therefore, their biometric properties, we used recordings containing heart sounds relating to the following pathologies: 1) innocent systolic murmur; 2) mitral regurgitation variations; 3) mitral regurgitation; 4) mitral stenosis; and 5) third heart sound.

Innocent systolic murmurs are not accompanied by other abnormal findings. One example is Still's murmur in children.

Mitral regurgitation is a condition in which disease or injury has caused the heart's mitral valve to become leaky. The four heart valves function as one-way valves, allowing blood to be pumped forward and preventing blood from regurgitating backwards. When the mitral valve becomes leaky, blood may back up into the lungs, causing a shortness of breath. The heart sound with mitral regurgitation is characterized by a "whooshing" sound (heart murmur) in the chest. The heart murmur is caused by the turbulent flow of blood from the left ventricle across the mitral valve and back into the left atrium.

Mitral stenosis is a narrowing or blockage of the opening of the mitral valve, which separates the upper and lower chambers on the left side of the heart. This prevents proper blood flow from moving between the left atrium (upper chamber of the heart) and ventricle (lower chamber of the heart). The third

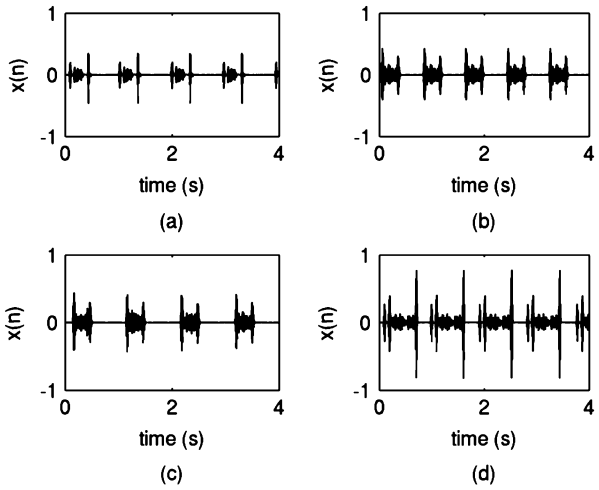


Fig. 1. Heart sound for (a) innocent systolic murmur, (b) mitral regurgitation variations, (c) mitral regurgitation, and (d) mitral stenosis.

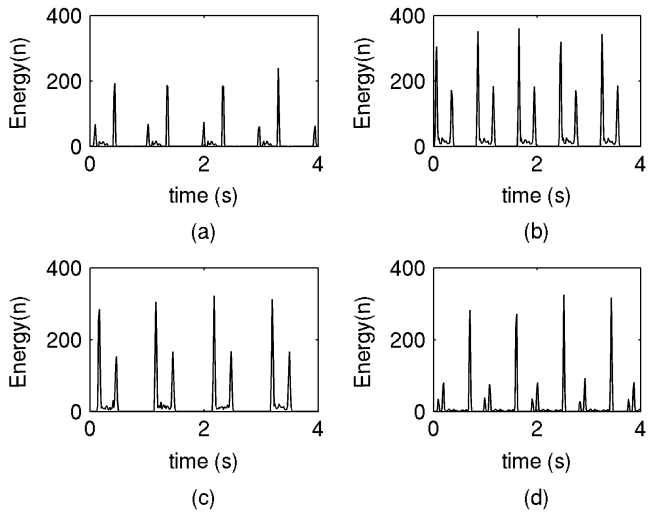


Fig. 2. Energy profile for (a) innocent systolic murmur, (b) mitral regurgitation variations, (c) mitral regurgitation, and (d) mitral stenosis.

heart sound or protodiastolic sound is not of a valvular origin, as it occurs at the beginning of the diastole just after S2. This sound occurs when the left ventricle is not very compliant, and at the beginning of the diastole, the rush of blood into the left ventricle causes vibration of the valve leaflets and the chordae tendinae. The health-care provider will listen to the heart and lungs with a stethoscope. A distinctive murmur, snap, or other abnormal heart sound may be heard. The typical murmur is a "rumbling apical diastolic murmur with pre-systolic accentuation." This means a rumbling sound is heard over the heart during the resting phase of the heartbeat. The sound gets louder just before the heart begins to contract.

For the subsequent analysis, the files were converted from the original MP3 format to a uniform 16 bits quantization format with a sampling rate of 11 025 Hz.

C. Cardiac Sounds Detection

In this paper, human identification is performed by analyzing the frequency characteristics of the sounds S1 and S2 in digital PCG sequences. It was thus necessary to implement a mechanism to identify the bounds of these sounds, in terms of samples, in an audio trace from cardiac auscultation.

A pathology-independent recognition system has to be capable of separating sounds from murmurs, as the latter are strictly linked to any pathology an individual may be affected by. Fig. 1 illustrates the trend of cardiac sounds with different kinds of pathology. As can be seen, it is possible to discriminate visually between sounds and murmurs. Cardiac sounds can be considered stationary if observed over short periods of time. To estimate the power spectral density, it is necessary to divide the signal into frames in which the sound remains sufficiently stationary. Following an empirical analysis, which took the typical duration of S1 and S2 sounds into account, a frame length of 20 ms was chosen. To blur the effect of sudden variations in the signal spectrum when passing from one frame to another, a partial overlap of 5 ms was introduced between samples belonging to adjacent windows: the frame extraction period is thus 15 ms. Extracting signal energy information every 15 ms also allowed us to obtain a more detailed profile

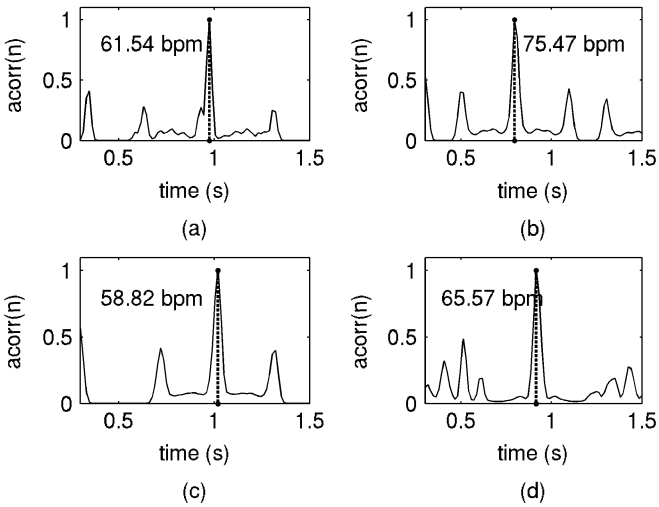


Fig. 3. Autocorrelation of energy profile and time period for (a) innocent systolic murmur, (b) mitral regurgitation variations, (c) mitral regurgitation, and (d) mitral stenosis.

of the energy trend [8]. Fig. 2 shows the energy profile of the single frames for cardiac sounds with various pathologies.

The first step consists of determining which sound—S1 or S2—has the greater energy; we will call this sound SX1. The first step in the mechanism to detect SX1 sounds is the determination of the frame with the highest energy value. We will call the index of this frame $I(0)_{\mathrm{MAX-SX1}}$. Then, the autocorrelation function of the energy signal is calculated. From analysis of this function, it is possible to determine the various periodic components of the energy signal and, thus, the basic period of the heartbeat. The periodic component with the highest energy will correspond to the position of the absolute maximum of the autocorrelation function. As the energy signal also has components close to continuous frequency, it is necessary not to consider low-frequency components. The absolute maximum was therefore not calculated over the whole interval but began from the value following the first relative minimum. Fig. 3 illustrates the autocorrelation function of the energy signal for the various cardiac sounds considered. The dashed vertical line corresponds to

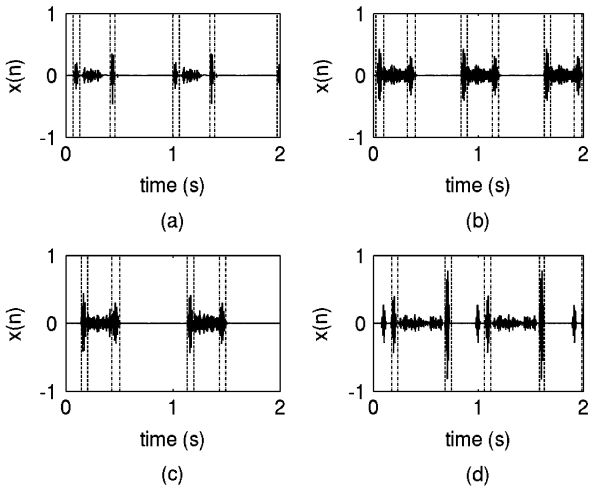


Fig. 4. S1 and S2 detection for (a) innocent systolic murmur, (b) mitral regurgitation variations, (c) mitral regurgitation, and (d) mitral stenosis.

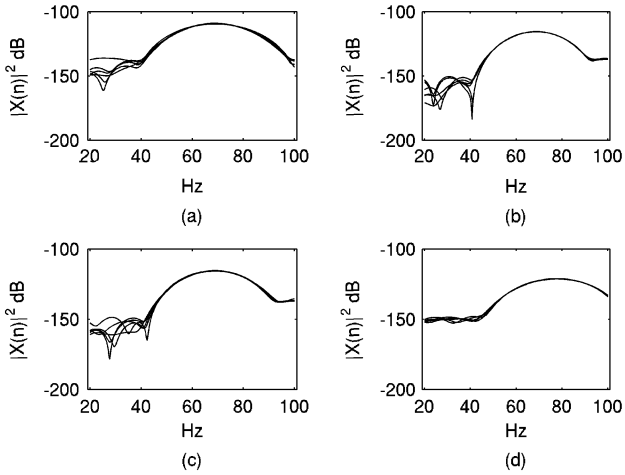


Fig. 5. S1 spectra relating to four different people.

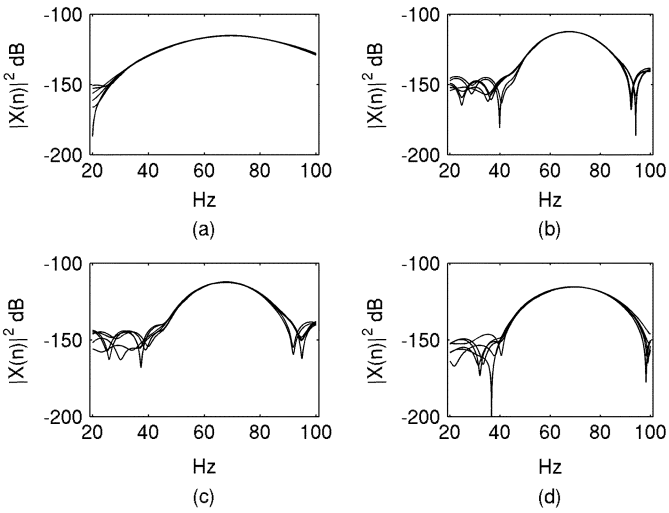


Fig. 6. S2 spectra relating to four different people.

the basic period of the heartbeat. Having determined the heartbeat period P in terms of the number of frames, it could be assumed that all of the remaining indexes corresponding to the maximum energy frames in SX1 sounds fall within the values $I(0)_{\mathrm{MAX-SX1}} \pm K \cdot P$ where K is a positive integer. In reality, this does not happen due to possible variations in the periodicity of the heartbeat. To follow these variations, the algorithm implements a step-by-step search for the position of the maxima. Having determined the frame $I(0)_{\mathrm{MAX-SX1}}$, the algorithm searches for all frames containing maxima to the right of $I(0)_{\mathrm{MAX-SX1}}$ and then all frames containing maxima to the left of $I(0)_{\mathrm{MAX-SX1}}$. It sets $\tilde{I}(-1)_{\mathrm{MAX-SX1}} = I(0)_{\mathrm{MAX-SX1}} - P$, for example, and then the value is corrected, if necessary, by determining the relative maximum around $\tilde{I}(-1)_{\mathrm{MAX-SX1}} \pm 3$ points. The new position will be the correct one $I(-1)_{\mathrm{MAX-S1}}$. At this point, the search for $I(-2)_{\mathrm{MAX-SX1}}$ starts from the new value of $I(-1)_{\mathrm{MAX-SX1}}$. The algorithm is iterated until all of the maxima to the left of $I(0)_{\mathrm{MAX-SX1}}$ are found. The procedure to find the maxima to the right of $I(0)_{\mathrm{MAX-SX1}}$ is the same, adding the periodicity value P found. Once the frames corresponding to the maximum energy values in all of the SX1 sounds have been determined, the start and end points are established, corresponding to the frames to the left and right, respectively, whose energy is 10% below the corresponding maximum value. The start and end positions of SX2 sounds are determined in the same way after having set all samples in the SX1 sounds to zero. On the basis of the relative distances between the SX1 and SX2 sounds, it is possible to establish whether SX1 is to be associated with S1 and SX2 with S2 or vice-versa. Fig. 4 shows an example of how S1 and S2 sounds are determined for the different cardiac sounds considered. The algorithm detects all of the S1 and S2 sounds present in the PCG sequence considered (generally comprising about six cardiac cycles).

D. Cardiac Sound Analysis in the Frequency Domain

Having determined the time intervals of the S1 and S2 signals, the signals are transformed from the time to the frequency domain [9], [10]. Experimental analysis of the duration of S1 and S2 signals showed that they never last more than 100 ms. To perform frequency analysis, each extracted signal was positioned at the center of a 100-ms window, setting any superfluous samples to the left or right to zero. To minimize the signal discontinuities at the beginning and end of each frame, the frame values were multiplied by a Hamming window. The energy of the S1 and S2 signals is essentially concentrated around frequencies below 200 Hz. To perform frequency analysis, the z-chirp (CZT) transform algorithm was used, calculating its values for frequencies between 20 and 100 Hz [10]. The value of the CZT samples was then normalized with respect to the sum of the square modulus of the values in each frame. The level of energy was thus obtained by transforming the value for each sample into decibels. Fig. 5 shows the signal energy spectrum for various repetitions of the S1 signal relating to cardiac sounds recorded from four different individuals. Fig. 6 shows the signal energy spectrum for various repetitions of the S2 signal relating to cardiac sounds recorded from four different individuals. Analysis of Figs. 5 and 6 shows that both S1 and S2 sounds for the same person essentially exhibit the same frequency spectrum in the interval considered.

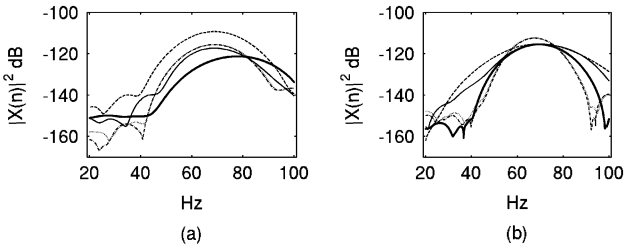


Fig. 7. Heart sound mean spectra relating to five different people. (a) S1 sound. (b) S2 sound.

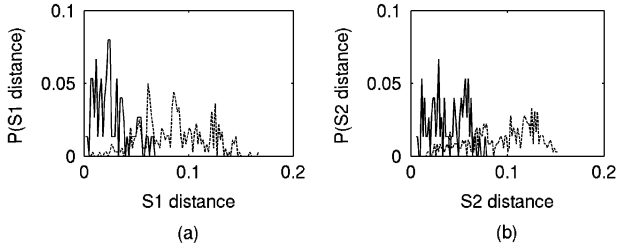


Fig. 8. Intraperson distance (solid line) and interperson distance (dashed line) distributions for (a) S1. (b) S2.

The possibility of distinguishing between individuals by using the spectrum of S1 and S2 sounds must, however, be supported by the fact that different trends are exhibited in recordings made from different individuals. To check this possibility, we calculated the mean signal energy spectra for five different individuals. Fig. 7 shows these mean spectra. As can be seen, it is possible, at least in some frequency intervals, to observe substantial differences between the curves.

E. Matching Algorithm

The last phase in the implementation of an automatic biometric identification system matches an (unknown) cardiac print to be analyzed with the templates stored in the system. Each template will comprise the spectrum, determined as described in Section III-D, of the S1 and S2 heart sounds relating to a certain number of heartbeats (in our tests, six heartbeats). The metric used to measure the distance between the two signal spectra was Euclidean. Considering the spectra as N-dimensional vectors, we have

$$d(X,Y) = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(X_i - Y_i)^2}$$

where X and Y are the vectors containing samples of the signal spectrum whose distance is to be measured, and N is the number of samples calculated for each frame.

When S1 and S2 signals are extracted from a cardiac sound recorded from the same individual, the distances with respect to the spectra are expected to yield lower values than those obtained when S1 and S2 spectra extracted from sounds recorded from different individuals are used. Following a series of recordings, we determined the distributions of the distance variable with S1 recordings made from both the same person and different individuals [Fig. 8(a)]. The same was done for S2 recordings [Fig. 8(b)]. In both figures, the solid line indicates the intraperson distance (i.e., the distance obtained using spectra from the same individual), while the dashed line indicates the interperson distance (i.e., obtained using spectra from different individuals).

F. Test Results on Cardiosource Database

The trend of the distance distributions was evaluated considering cardiac sound recordings from five different individuals, as described in Section III-B. Six different heartbeats were extracted from each recording, thus giving six S1 and six S2 sounds. The number of distances for each type of sound, calculated from recordings made on a single person, will be 75, having $15(5+4+3+2+1)$ pairs of distinct sounds for each person. The number of distances for each type of sound, calculated from recordings made by different individuals, will be 360, having 36 $(6{\cdot}6)$ pairs of sounds (for each pair of individuals) and ten pairs of individuals $(4+3+2+1)$. The analysis of Fig. 8(a) and (b) shows that the hypothesis regarding the probability of finding greater distances when cardiac sounds from different individuals are considered is confirmed. It may be appropriate in an identity test to consider the distances calculated on both S1 and S2 sounds. Using bimodal (or multimodal) distribution, the FRR and FAR values generally decrease. We can consider a two-dimensional vector in which the first component is the S1 distance and the second is the S2 distance. On a Cartesian plane, the points relating to distances calculated from recordings made on the same person would be expected to concentrate in the proximity of the point $(0,0)$, whereas those relating to distances calculated from recordings made on different individuals would be expected to be quite far from the point $(0,0)$. The statistical analysis of various recordings will make it possible to determine a curve that minimizes the misclassification error (i.e., it minimizes the misidentification probability once the false reject probability is fixed, and vice-versa).

When recordings are available containing different heartbeats for a single individual, in order to minimize errors linked to signal processing operations, it is suitable to perform a separation analysis on the basis of the mean distances calculated between both sounds belonging to the same person and sounds belonging to different individuals. In practice, if we indicate the ith repetition of the S1 spectrum for person X as X_i^{S1} and the ith repetition of the S1 spectrum for person Y as Y_i^{S1}, we can define

$$\tilde{d}^{S1}(X,X) = \frac{1}{N_X - 1}\sum_{i\neq j}^{N_X} d\left(X_i^{S1}, X_j^{S1}\right)$$

$$\tilde{d}^{S1}(X,Y) = \frac{1}{N_X \cdot N_Y}\sum_{i,j}^{N_X,N_Y} d\left(X_i^{S1}, Y_j^{S1}\right).$$

In the same way, we can define the distances for S2 sounds to obtain the second coordinate of the mean point. We thus obtain the pairs of points $(\tilde{d}^{S1}(X,X), \tilde{d}^{S2}(X,X))$ for all of the individuals X considered that describe on the Cartesian plane the distances relating to the same person and $(\tilde{d}^{S1}(X,Y), \tilde{d}^{S2}(X,Y))$ for all possible pairs of individuals (X,Y) that describe the distances relating to different individuals.

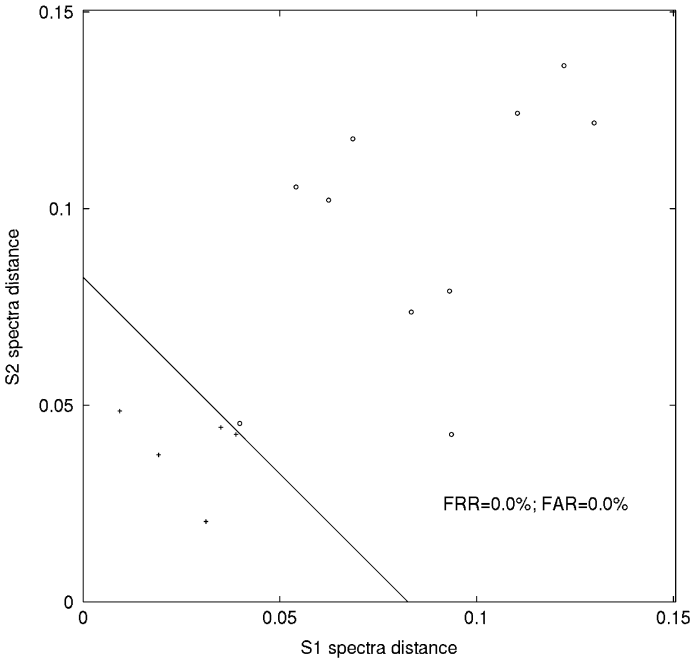


Fig. 9. Distribution of points relating to S1 and S2 distances: intraperson (plus sign) and interperson (circle).

In this way, considering a population of five individuals, it is possible to determine five different points which describe the pairs of mean distances calculated from recordings made on the same person and ten $(4+3+2+1)$ points that describe the pairs of mean distances calculated from recordings made on different individuals. Fig. 9 gives the distribution of the mean points on the Cartesian plane, a plus sign indicates the points representing the distances calculated from recordings on the same person and a circle indicates those representing distances calculated on recordings from different individuals. As can be seen in this case, bearing in mind the relatively small size of the database, it is always possible to distinguish between individuals via analysis of cardiac sounds, obtaining a null value for both performance indexes—FRR and FAR.

G. Robustness to White Gaussian Noise

In an initial test of robustness of the algorithms for the extraction of heart sounds to be used for identification purposes, a series of tests was performed adding white Gaussian noise to the five recordings so as to obtain mean signal-to-noise ratios (SNRs) of -5, 0, 5, 10, 15, and 20 dB. Fig. 10 gives an example of heart sound identification relating to the "Third Heart Sound" recording in the Clean case and in the presence of the various noise levels considered (i.e., SNRs of -5, 5, and 15 dB).

As can be observed, up to $\mathrm{SNR} = 5$ dB, the algorithm for the detection of S1 and S2 sounds continues to segment correctly, as in the Clean case. It is only with very low SNRs that the segmentation system is affected by the presence of noise and segments a wider portion of the signal. Figs. 11 and 12, which refer, respectively, to S1 and S2, show the mean spectrum for each person, calculated in the two extreme cases: SNRs of 20 and -5 dB. As can be seen from Figs. 11 and 12, there are no substantial variations in the spectrum due to the addition of white noise at the frequencies considered, which range from 20 to 100 Hz. From Fig. 13, it is possible to deduce that there are no considerable degradations, not even in terms of the degree of separability between the classes linked to the identification process and in the presence of a large amount of noise.

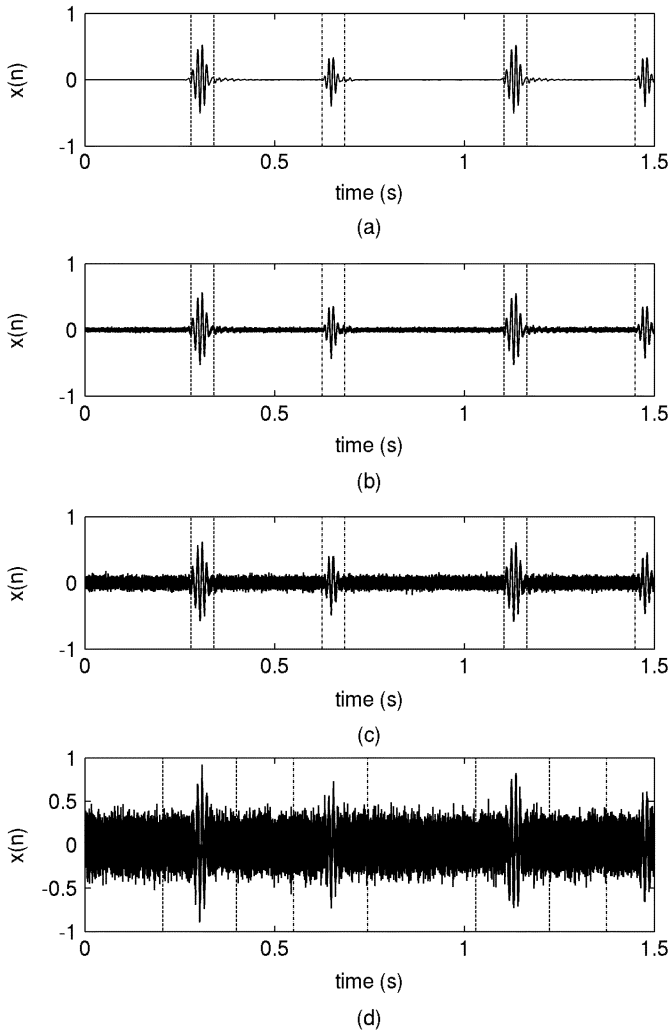


Fig. 10. S1 and S2 detection of heart sounds in the presence of Gaussian white noise. (a) Clean. (b) SNR = 15 dB. (c) SNR = 5 dB. (d) SNR = -5 dB.

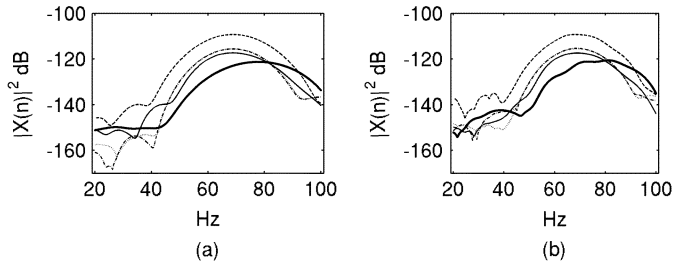


Fig. 11. S1 spectra relating to five different people at different SNRs. (a) 20 dB. (b) -5 dB.

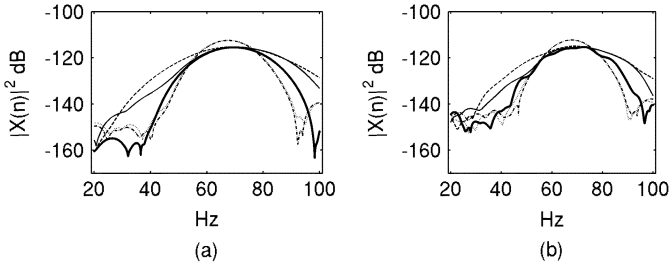


Fig. 12. S2 spectra relating to five different persons at different SNRs. (a) 20 dB. (b) −5 dB.

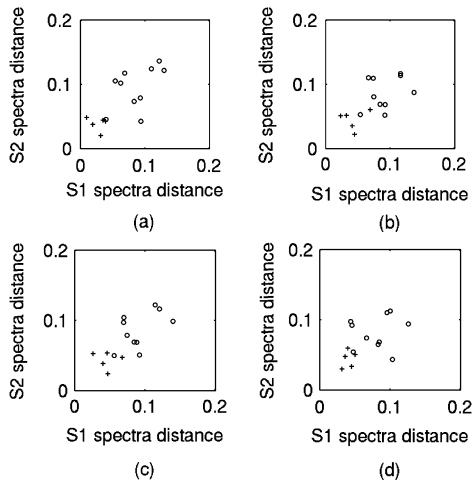


Fig. 13. Distribution of points relating to S1 and S2 distances: intraperson (plus sign) and interperson (circle) at different SNRs. (a) Clean. (b) 15 dB. (c) 5 dB. (d) −5 dB.

IV. Test Results on Recorded PCG Sequences

A. Database

In an initial test of the potential of the segmentation and biometric identification algorithms proposed in Section III, when applied in a context in which the PCG recordings are made by nonexperts, we created a database containing heart sound recordings from 20 different people (twelve females and eight males). The recordings were made using a Littmann 4100 WS electronic stethoscope three different times in such a way that the interval between recordings on any one person was no less than two days over a two-month period. Compared with the sequences for the Cardiosource database, the recordings made using the electronic stethoscope were noisier, as shown by the example given in Fig. 14. This degradation is, in part, due to inexpert use of the stethoscope (applying different amounts of pressure when positioning and moving the stethoscope during recordings) and to the presence of additional noise due, for example, to movements and/or breathing by the subject being examined. In this respect, however, the database is a more faithful reproduction of the conditions typically encountered in a real application context. For a subset of people, the heart sound acquisition process was performed by placing the stethoscope over four different typical auscultation regions (Fig. 15): aortic (A), between the second and third intercostal spaces at the right sternal border; mitral (M), near the apex of the heart between the fifth and sixth intercostal spaces in the mid-clavicular line; pulmonic (P), between the second and third intercostal spaces at the left sternal border; and tricuspid

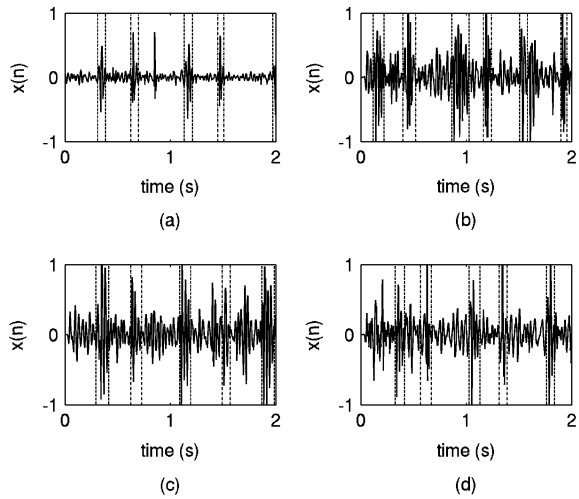


Fig. 14. Examples of PCG signals acquired using the Littmann 4100WS electronic stethoscope.

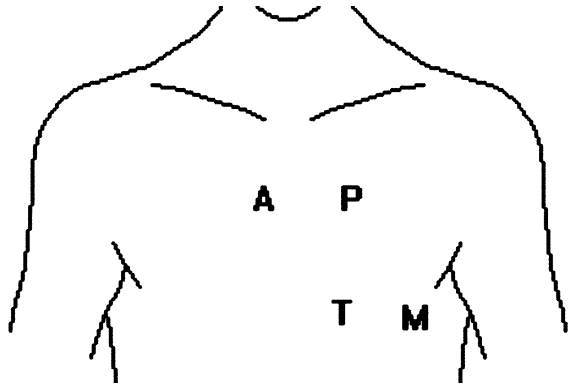


Fig. 15. Typical auscultation regions.

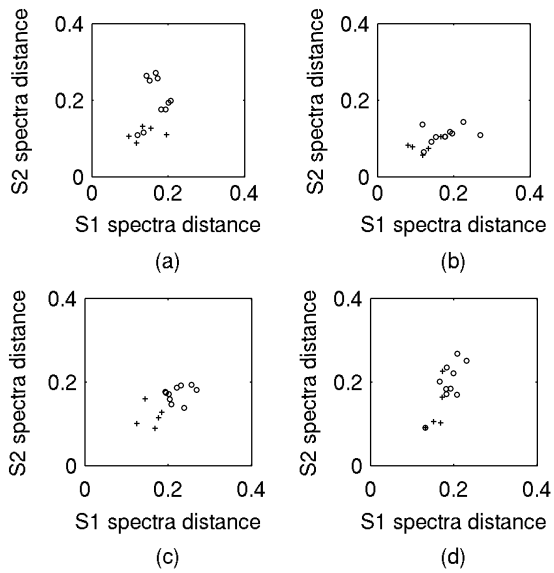


Fig. 16. Distribution of points relating to S1 and S2 distances: intraperson (plus sign) and interperson (circle) placing the stethoscope over the (a) aortic region, (b) mitral region, (c) pulmonic region, and (d) tricuspid region.

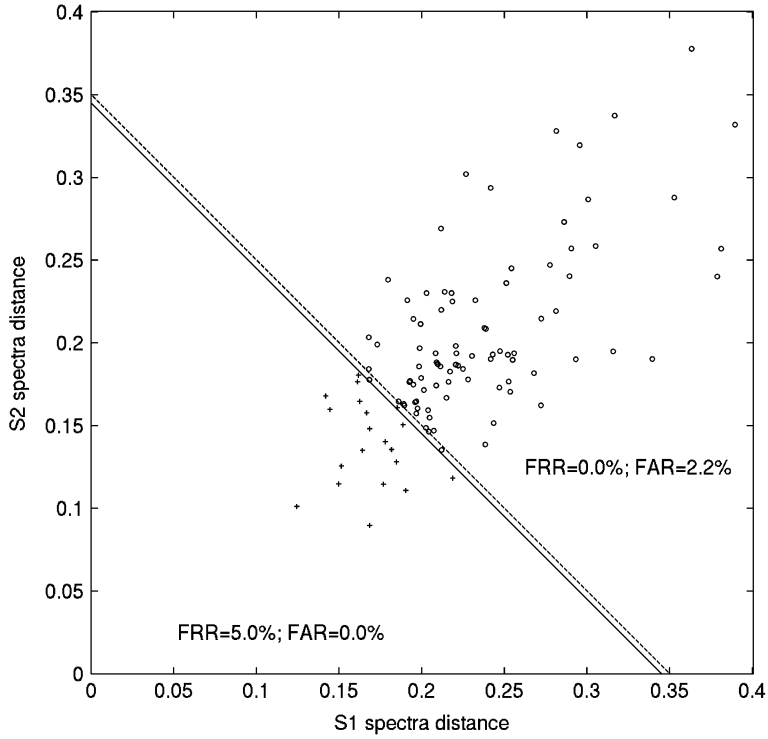


Fig. 17. Distribution of points relating to S1 and S2 distances: intraperson (plus sign) and interperson (circle) using heart recordings for 20 people recorded via the Littmann 4100 WS stethoscope.

(T), between the third, fourth, fifth, and sixth intercostal spaces at the left sternal border. This was done to try to identify the position which gave the best performance in terms of the degree of separability between the intra and interperson difference. All of the original recordings were converted into uniform 16 bits quantized audio files with a sampling rate of 11 025 Hz.

B. Results of Identification Phase

The performance of an identification system based on frequency analysis of heart sounds depends on both the segmentation of S1 and S2 sounds and the subsequent matching phase. In regards the segmentation phase, the tests performed on the database of recorded PCG sequences match the performance obtained in tests carried out on the cardiosource database. Fig. 14(a)–(d) shows the correct behavior of the S1 and S2 detection algorithm in relation to examples of PCG sequences. The segmentation algorithm, which is essential for the subsequent matching phase, generally segments the sequences analyzed correctly and so is very robust to the degradation typically occurring in PCG sequences recorded in a real application context. As far as the matching phase is concerned, we first conducted a preliminary study to identify the heart auscultation region (i.e., one of the four auscultation positions for the stethoscope), which gives the best degree of separability between classes for intra and interperson distances. PCG recording sessions were run on five different people, placing the stethoscope over the four auscultation regions. Fig. 16 illustrates the results of the preliminary investigation. Clearly, the recordings from the aortic and pulmonic regions exhibit a greater degree of separability between the classes. It was then decided to extend the database, this time recording only PCG sequences relating to the pulmonary valve, and obtaining a total of 20 different groups of PCG sequences for 20 different people (twelve females and eight males). Using this database, the performance of the identification system was evaluated in terms of FRR and FAR. The results obtained are shown in Fig. 17 which, like Fig. 8, illustrates the points relating to intraperson distances (plus sign) and interperson distances (circles) obtained from recordings made on the pulmonary valve alone. Although the recordings are considerably affected by noise of various kinds, and although the stethoscope was not positioned by specialized medical personnel, the results still show a good degree of separability between the intraperson and interperson distances. Analyzing the S1–S2 distances, it is possible to identify two lines with a negative unit slope which, respectively, give the extreme conditions of null false accept rate (FAR $= 0\%$) and null false reject rate (FRR $= 0\%$). A person is correctly declaring his or her identity if the matching process yields a point below the first line; if the distance is above the second line, the system detects the person as lying about their identity. The values between the two lines indicate the impossibility of verifying the person's identity, thus requiring further investigation based on other biometric parameters. This range includes both people who declared their true identity but whom the system rejects with an error rate of FRR $= 5.0\%$, and people whose false identity is not recognized by the system with an error rate of FAR $= 2.2\%$.

V. Conclusion

This paper has proposed a preliminary study for the implementation of an automatic identification system based on frequency analysis of the main heart sounds S1 and S2 recorded

by an electronic stethoscope. The first part of this paper presents an algorithm to detect heart sounds, which is essential for the subsequent frequency analysis and signal matching phases. Analysis of the inter and intraperson distances measured for S1 and S2 indicates that there is a good degree of separability between the two sets of values, in particular when PCG sequence recordings are made, placing the stethoscope over the pulmonary valve. The results of the tests performed indicate that the two main heart sounds in a PCG sequence are thus a good physiological sign that can be taken into consideration in modern biometric systems.

REFERENCES

[1] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: A tool for information security," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 125–143, Jun. 2006.

[2] A. K. Jain, R. Bolle, and S. Pankanti, *Biometrics: Personal Identification in Networked Society*. Norwell, MA: Kluwer, 1999.

[3] A. K. Jain, L. Hong, and S. Pankanti, "Biometrics: Promising frontiers for emerging identification market," in *Commun. ACM*, Feb. 2000, pp. 91–98.

[4] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security & privacy concerns," *IEEE Sec. Privacy Mag.*, vol. 1, no. 2, pp. 33–42, Mar./Apr. 2003.

[5] V. Nigam and R. Priemer, "Cardiac sound separation," *Comput. Cardiol.*, pp. 497–500, 2004.

[6] "Heart songs" [Online]. Available: http://www.cardiosource.com/heartsounds/index.asp.

[7] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol., Special Issue Image- and Video-Based Biomet.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

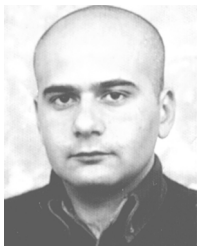
[8] A. Haghighi Mood and A. N. Torry, "Coherence analysis of multichannel heart sound recording," *Comput. Cardiology*, pp. 377–380, 1996.

[9] A. Djebbari and B. Reguig, "Short-time fourier transform analysis of the phonocardiogram signal," in *Proc. IEEE Int. Conf. Electron., Circuits, Syst.*, 2000, pp. 844–847.

[10] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

Francesco Beritelli (M'95) received the Laurea degree in electronic engineering and the Ph.D. degree in electronics, computer science, and telecommunications engineering from the University of Catania, Catania, Italy, in 1993 and in 1997, respectively.

Currently, he is Assistant Professor in the Department of Computer Science and Telecommunications Engineering at the University of Catania. His main research activities are in the area of robust speech signal classification and recognition, variable bit-rate speech coding, and adaptive-rate voice transmission for IP telephony applications. His interests also include the field of biometric identification and cardiac signal processing.



Salvatore Serrano received the Ph.D. degree in computer science and telecommunications engineering from the University of Catania, Catania, Italy, in 2003.

Currently, he is a Fixed-Term Researcher in the Computer Science and Telecommunications Department at the University of Catania, Enna, Italy. His research activities are in the area of speech signal processing: speech coding and recognition, speaker recognition/verification, human emotion classification by speech analysis, biometric identification, and voice transmission over IP.